DOCUMENT RESUME

ED 048 845 JC 710 081

AUTHOR Gold, Ben K.
TITLE Evaluation of Programs.
PUB DATE [71]
NOTE 20p.; Paper presented at a conference sponsored by the Compensatory Education Project, Coordinating Board, Texas College and University System, April 5-6, 1971, Austin, Texas

EDRS PRICE EDRS Price MF-$0.65 HC-$3.29
DESCRIPTORS *Evaluation, *Evaluation Criteria, Evaluation Methods, *Evaluation Techniques, Inspection, *Junior Colleges, Measurement, *Program Evaluation

ABSTRACT

Evaluation in its true concept should be a process for collecting information to make better decisions. The author discusses in detail four planning stages to evaluate programs. The first stage of the process is to ascertain the decision areas of concern. In the second stage, the evaluator must select the appropriate information-gathering instruments. In the third stage, the data must be collected and analyzed in advance of the decision maker's deadline. The final stage is to report the findings to the decision maker, in time for him to use them, and in a form he can understand. The author offers eight references on the subject of evaluation. (CA)

ED048845

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY

EVALUATION OF PROGRAMS

Ben K. Gold
Director of Research
Los Angeles City College

REACHING FOR THE IDEAL:
Serving the Disadvantaged Through the Community College

A Conference sponsored by the Compensatory Education Project, Coordinating Board, Texas College and University System, April 5-6, 1971, Sheraton-Crest Inn, Austin, Texas.

JC 710 081

UNIVERSITY OF CALIF.
LOS ANGELES

APR 26 1971

CLEARINGHOUSE FOR
JUNIOR COLLEGE
INFORMATION

EVALUATION OF PROGRAMS

No one needs to be told today that the educational world, as well as the world in general, is in a state of turmoil. The quiet solitude of academia is the thing of the past, if it ever really existed. Society is looking more and more to education to offer some rays of hope in the darkness of problems that loom larger and larger and despair of solution. Education is responding with a bounty of new ideas, methods, and programs. The educational institution is rare that is not in torment over its very place and purpose in today's complex world.

The junior college in many respects is the most tormented segment of education. Reasons for this are manifold: its open door policy, the enormous range of talents, interests, and backgrounds of its students, its multitude of course and curricula offerings -- just to name a few. Concomitantly, and for these very reasons, the junior college is looked to by many to offer great hope for finding at least some partial solutions to the problems now being faced. Witness the amounts of money the urban colleges are obtaining through governmental and private grants. Witness this conference today.

What I am here to discuss today is not the problems, not the grantsmanship, not the specifics of the programs being tried, not the reports required by the funding agencies, although all these things are related to our discussion in some way. I am here simply to make a plea that you don't wait until your program is over to decide whether or not it's worth anything, but that you start, right from the initial planning stages, to -- if you will -- evaluate your programs.

Evaluation is a popular word today. It is on the minds of many people. We hear talk of revision of grading procedures of students -- the no "F" policy; we hear talk of students wanting to evaluate teachers; we are all, at least in California, concerned about attacks on our tenure system; we read requirements written into grants; and there is that word that is beginning to haunt everybody -- you saw it on the cover of the March Junior College Journal in two inch headlines -- "Accountability." Yes, evaluation is on people's minds. As we discuss it today, lets begin where we should begin, with a definition of terms.

Just what do we mean by evaluation? My Merriam-Webster tells me it means "finding the value of," or "appraisal." Very little help here. I suggest you think for a moment about what the word evaluation means to you in an educational context. I suspect that you would equate it with one or more of the following ideas: possibly observation, measurement or testing of some kind; equating actual performance with expected or hoped for performance; or possibly some sort of professional judgment I would concur that all these are aspects of evaluation but I submit that none is adequate for a usable definition of evaluation that we can apply to college programs. The definition of evaluation that I would like to concentrate on today is one phrased by the people in the Teaching Research Division of the Oregon State System of higher education. Here it is:

> <u>Evaluation is a process of examining certain objects and events in the light of specified value standards for the purpose of making adaptive decisions</u>.

Note right away some obvious principles that follow from this definition. First, evaluation is a process of gathering information; second, the information collected will be aimed toward its use in aiding the decision maker; third, information must be presented to the decision maker in a form that he can use it effectively; and fourth, different kinds of decisions may require different kinds of evaluation procedures.

As viewed by the Center for the Study of Evaluation at UCLA, the process of evaluation consists of four stages. Their definition includes these four stages, as follows:

> Evaluation is the process of ascertaining the decision areas of concern, selecting appropriate information, and collecting and analyzing information in order to report summary data useful to decision-makers in selecting among alternatives.

You will observe that the two definitions are quite similar.

The key concept in the definition of evaluation I am proposing for your consideration today is that the evaluator's function is to provide to decision makers information that can be used effectively to make decisions about alternative courses of action.

Let me digress for a moment to comment on evaluation as opposed to research. Certainly research techniques are employed in the evaluation process, but to me the key difference is the concept of the value standard. To a researcher, the prime concern is a functional relationship -- to discover or explain some phenomena. This usually means he will design rather comprehensively his plan of action. The evaluator on the other hand is concerned that better decisions will be made and he may revise his plan considerably as the project progresses.

Let us consider now the first stage in the evaluation process -- that of ascertaining the decision areas of concern. Thinking about this raises questions such as the following: What is the purpose of the evaluation?

Who will make the decisions? What criteria will be used by the decision maker? What are the value standards he will use? Just what is the program supposed to accomplish? Who decides what the program is supposed to accomplish? And quite a few more that I am sure you could think of.

All of this takes us back ultimately to the philosophical principles on which the college operates -- what is the nature of the good life? -- what is important? -- principles such as: "everyone should have the opportunity to become educated to the maximum of his capabilities and interests;" or, "the college should maintain an environment conducive to the development of programs which respond to the needs of students in a changing society." For more such principles read the opening pages of any college catalog. Or see the set of recommendations so eloquently prescribed by the staff of the Compensatory Education Project.

These principles suggest the kinds of behavior patterns, the types of values and ideals, and the habits and practices that the program will be aimed at, and from these philosophical principles are derived the goals of the program, usually in pretty general terms. These goals will guide the choice of activities to be included in the program and from these goals should flow the specific objectives of the program, which in a real sense are the operational definitions of the goals. It is certainly to be hoped that this hierarchy leading to specific objectives is sensitive to the society the college serves, to the student to whom it directs its efforts, and to the disciplines involved. By this I mean clearly to recommend that in planning the program the community, the students and the faculty should have a voice. Let me suggest, in addition, that the program evaluator be included in these initial planning conferences, primarily to make sure that the objectives for the program finally agreed upon will be stated in a form amenable to evaluation.

Let me also suggest that in these early planning stages you include someone knowledgeable in the area of electronic data processing, primarily to make sure that the information to be collected will be collected in a form that will expedite its analysis.

It is difficult to overemphasize the importance of good well-stated objectives. Some of the properties of a good well-stated objective are: (1) it should be defined clearly enough so that all involved in the program can recognize and understand it; (2) the activities necessary to its fulfillment are possible; (3) there should be serious intent to achieve it, even at considerable cost; and (4) there must be some way of determining, or at least estimating, the degree to which the objective is actually realized. This last point is probably the most important and at the same time the most difficult to accomplish.

On the subject of statement of objectives, let me recommend to you Robert Mager's delightful little book on Preparing Instructional Objectives. If you haven't read it, you will find it well worth your reading, not only for your clasroom work but for preparing objectives for your programs.

Another useful device is one attributed to C. F. Paulson of Oregon called the A B C D's of good objectives: A, you should consider the audience, describe your learners, what are their entry characteristics into the program; B, behavior, what is the learner expected to do; C, conditions, what circumstances, givens and props provided, and D, degree, what is the criterion or by what do you determine whether or not the objectives have been met. Mager says it beautifully: "You should be able to find some way to evaluate anything you think important enough to spend a significant amount of time teaching. If you find something you feel sure you cannot measure the place to put effort is in trying to find some way to measure it." And, we might add, to be sure you know what it is you are

trying to measure, give considerable thought to the statement of your objectives.

Now, before we leave this first task of an evaluator, that of ascertaining the decision areas of concern, we must include another aspect. In ascertaining what the decision making body is attempting to do, the evaluator must know something of his theater of operation. In addition to identifying the outcomes or objectives he must also obtain an adequate description of the population to be studied and the criteria for their selection and an accurate description of staff, media, facilities, and planned activities. We of course keep in mind that many of these things will be preliminary and subject to change, but the evaluator should have just as much information as possible before he starts out on the remaining stages of the evaluation process.

Let us now turn to the second area in which the evaluator must become involved, that of selecting the appropriate information. Now that we have some idea about what the program is trying to accomplish and we know something of the situational factors, the next question is "what kind of information do we want to collect and what instruments will help us get it?"

There are several concerns here of course. First we might ask how much evaluation will be needed, or wanted, or will be able to be supplied. What kind of a budget is there for the evaluator? How much time will be allowed? How much help will he get? How much evaluation do the project directors really want? Do they want it badly enough to support it? Do they want what they really need? We must consider financial constraints -- how much money is available, how much of that money the administrators of the program will choose to spend on the evaluation. We may have to make some important decisions -- for example, deciding between finding out information

from written questionnaire or by personal interview. Personal interview is considerably more expensive than the written questionnaire but may provide considerably more information. There are situational factors to be considered also. We must know something about the respondents. We must know what kind of knowledge they have about particular topics if we are going to ask them questions on a questionnaire for instance. We must be able to know what amount of thought they will be able to give to these questionnaires. We must know something about their ability to communicate. Will they be able to answer the questions we are asking? There are many things to be considered before we just dive in asking questions or giving tests.

Also, when we are beginning to think about what kind of instruments we are going to use to elicit information, let's be careful not to just select the obvious ones and the ones easy to get. I certainly cannot deny that one has to consider this selection problem in the light of all the constraints on the evaluator, but maybe if you look around a little, you might be able to find something which won't cost you any more, which won't take any more of your time, but which will give you much more reliable and valid information. For example, its easy to settle for a grade point average as a measure of learning in a particular course but does it <u>really</u> give you an <u>accurate</u> measure of <u>learning</u>? Maybe it does, but I have my doubts in many instances. In any case, I'd like to consider some other type of measure, hopefully to get some kind of cross validation. As you'll see, this concept of cross validation is one I consider quite important. I think we should try as many approaches as possible. Don't settle for some pet idea or technique but consider other possible ways of looking at the situation, keeping in mind the objectives of the program.

Now, how do we go about finding instruments that will help us attain the measurement of our objectives?

There are many attributes of measuring instruments that are important, three in particular I think that we should look for in evaluating a program. First, the instrument must have reliability, meaning that whatever the instrument measures, it measures it consistently -- it can be depended upon. Secondly, it must have some kind of validity -- meaning it measures what it is supposed to measure -- it isn't measuring something else completely different from what we want. The third one is not always considered, but I think it is important -- especially when we are considering programs for the disadvantaged; the instruments should have some kind of relevance. This word is abused and overused today, but what I mean is the instrument should not be an affront in any way to the persons who are going to be asked to respond to it.

Let us turn now to specifics: how do you find instruments?

I suggest first of all that you take a look at what standardized tests are available. There are many advantages to using standardized tests. They have already been checked for reliability and validity, they're generally less expensive and you can usually find a critical review in the Mental Measurements Yearbook, or elsewhere, in which some outside person tells you what he thinks of the test. There are disadvantages of course. One of the major ones is that they are normed on groups which are generally not the same type of group that you are working with. And the validity coefficient they give you may be related to a criterion which is not one you are concerned with. Also there are administrative problems -- you must be sure to administer the tests under conditions specified by the test publisher. But let me suggest that you at least look to see if you can find one. Now I think many times you are not going to be able to find one, so I am going to suggest some types of things you might do to make up one of your own.

Especially when you are trying to measure objectives in the affective area -- where you want to learn something about attitudes of people -- very seldom will you find a standardized instrument that is entirely appropriate to your specific situation. So why not try making one up?

Let me mention three of the more popular and I think useful ways of constructing a home made attitude scale.

First, consider the one that is commonly known as the Thurstone scale. Here's generally the way one goes about it. You write some statements about whatever you are trying to get an attitude toward, making some of the statements very favorable, some of them very unfavorable, and some neutral. I suggest you make forty or fifty of these statements. Then, give them to somebody, not just one somebody, but several -- as many as you can get, hopefully 20 or more. Try to get those types of people who will be similar to the ones to whom you are going to administer this instrument and ask them to rate each statement on a scale, usually chosen from zero to ten, as to whether or not they find it favorable or unfavorable. When this is done you find the median scale value for each student; that is, the middle value for all these judges and select items for your final instrument whose median scale values range as far as possible over the full scale. You may select 10, 15, 20, or so for your instrument. Next put them in a random order and then ask the respondent simply to check which ones he agrees with. His particular score on this instrument will be the median of the scale values for the items that he has checked, and you have a measure of his attitude toward the object under consideration. It's a very rough instrument, as all these home made things are, but it will give you some idea of an individual's attitude compared to those on which you have based what we might call norms, in other words, your judges' responses.

Another and probably more popular technique is the so called Likert scale. The Likert scale differs from the Thurstone in several respects. First, you write some statements, with about half favorable and half unfavorable. To each statement you attach a scale, usually of 5 to 7 points. On the 5 points, for example, you might use the words "strongly agree," "agree," "neutral or no opinion," "disagree," or "strongly disagree." After you have prepared your statements, you again give them to some judges. Hopefully this time you can find two groups, one which will react favorably and the other unfavorably. Then after you have selected on the basis of your judges' response the items which are favorable and unfavorable, you administer these to your group, scoring the favorable items with the values 1, 2, 3, 4, 5 from strongly agree to strongly disagree; on the unfavorable reverse the sequence to 5, 4, 3, 2, 1. Then you simply get a total score for the individual. The Likert scale as I mentioned is probably the most popular of this type of thing and is quite appropriate when you're interested in some kind of relative index to compare one person with another, or to compare a pretest and posttest administered to the individual at the beginning of the program and later on after he has been subjected to the treatment of the program.

Let me mention one other, the "semantic differential." This is a device using a set of bipolar adjectives; for instance weak-strong, good-bad, important-unimportant. The person is asked to check a point on some kind of graphic scale between these two extremes indicating his reaction to the particular object that you are concerned with. This semantic differential is a relatively quick method of obtaining measures for several different objects about which you want to measure attitudes. If you would like to go back to Osgood's original work on the semantic differential you can find some

particular sets of adjectives which will combine to give you a stronger measure of certain types of attitudes. If you do that however you are almost back to the concept of the standardized test. What I'm suggesting is that you make up your own, check it out with some people in advance and use it. For example, you might put "myself" at the top and ask the respondent to indicate his feelings on several scales. Take a look at the difference in feelings about different persons' self-appraisal. You might use "my instructors," "my text books," or a whole variety of concerns. Again, these are very rough measures and I'll have a difficult time defending them against an expert who might challenge them. Yet I believe they are useful if done carefully and cross validated with other types of measures.

To recap then, the Thurstone is a useful device to get some kind of an absolute measure of attitude. The Likert measure is a very popular one useful for relative indices comparing one person with another or pretest with posttest and the semantic differential is a quick method of obtaining measures for a variety of objects. In all of these written home made instruments let me caution you about one particular type of error that is very often not thought about but is easy to overcome if you do think about it -- that is the so-called "expert" error. You forget sometimes when you're writing the statements that you are familiar with what's going on, you are familiar with a certain lingo or jargon, or pedagese. Don't forget that maybe the person who is going to read it isn't. Make your statements and instruction in as basic English as you can. Be sure that the person who is going to read the instrument will be able to understand it.

You might think at this point, why all these fancy ways to measure attitudes, why not design a questionnaire and just ask the person. And why not? A questionnaire can often be a very useful device, but let me point out two or three things about questionnaires that I think should concern you. In

the first place, be sure you know what you're trying to find out. Don't just ask questions for the sake of asking questions. I think it is quite important that every question is put on a questionnaire with a certain object in mind. There are problems too with the collection and analysis of data, so think carefully about whether or not you're going to use direct short answers, multiple choice type questions or whether you're going to use open-end questions. Think ahead how you're going to categorize the information you will get with open-ended questions. It might be difficult. Then too, I would suggest that when you start to make out a questionnaire you sit down and write at least 5 times as many questions as you think you are going to use. Then divide these up and try them on some people preferably similar to the ones who will respond to your questionnaire. These might be students in the program, or the staff, or the community or whomever you can find. Information obtained from the questionnaire will be valuable if it has some object in mind and it can be reasonably obtained and is reliable. So check it out first, run some sort of a pilot and keep only those questions which you are sure everyone can understand and respond to. If you're going to use a mailed questionnaire for any reason, I suggest you consider the following 7 factors which have been known to affect the return.

1. Who sponsors it?
2. How attractive is it?
3. How long is it?
4. What kind of cover letter goes with it?
5. How easy is it to fill out and return?
6. What inducements are there for it to be returned?
7. What population do the respondents come from?

Most people will tell you that if you get 50% return from a mailed questionnaire that you have a very good response. 25 to 30% is average. I find personally in dealing with students, graduates, and other follow-ups of people who are involved in certain programs that we have no problems in getting 60% or thereabouts providing you make the instrument attractive,

short, and to the point, you enclose a self-addressed stamped return envelope and you entice them in your cover letter by point out that their responses will help future students. I know of one study which showed that use of attractive memorial stamps increased the response rate significantly.

What about interviews? Interviews can give you much more in-depth information than questionnaires. They are however, time-consuming and difficult to perform. This a lot of people don't seem to realize. They think all you need to do is go ask questions; but the order of the questions, the type of questions, the attitude of the interviewer, his voice inflection, the way he's dressed, the type of person he is, all can have a tremendous effect on the responses, so while I think interviewing is a very fine technique, I would advise you not to use it unless you think through carefully the various difficulties involved and read some of the literature on how to conduct a good interview.

Another type of measure that is very useful in some situations is the so-called non-reactive or unobtrusive measure. This is in a sense the spy approach to find out about people without their knowing about it. I don't like to put it in the context of spying, but there are some things you can look at without bothering people and without interfering with the program which will give you some useful information. For example in an Art exhibit on campus it was found which particular item of art was the most popular by simply looking at the wear on the rug.

There are three general sources for these unobtrusive measures, one of which is physical traces of past behavior as in the example which I mentioned. Another one is archival records where you can find historical data about students, and the third is plain and simple observation. We won't take time today to go into anymore particulars on this, but I might suggest a very nice

little book called Unobtrusive Measures written by Webb and several others that I suggest you look into if you are interested in this approach to obtaining information.

Professor W. B. Michael and N. S. Metfessel of U.S.C. have tabulated a long list of criterion measures that might have some usefulness in evaluating school programs. Not all of them have value in the junior college situation, but a good many of them do. Let me just read a partial listing in one of their five categories - this one listing "indicators of status or change in student behaviors other than those measured by tests, inventories, and observation scales in relation to the task of evaluating objectives of school programs":

absences, frequency of
anecdotal records
appointments, frequencies with which they are kept or broken
attendance, frequency and duration
books, numbers checked out of library, numbers reported read
changes in program, frequency of occurrence
choices expressed or carried out - vocational and educational
citations - commendatory in both formal and informal media of communication
contacts - frequency or duration of between observed person and significant other person
disciplinary actions taken - frequency and type
dropouts
elected positions
extra curricular activities
grade point average
leisure activities
library card - possessed or not
load - number of units
peer group participation
recommendations
referrals
skills - craft, P. E. and others not measured by available tests
transfers

I am sure that two or three people involved in a college program could sit down for a couple of hours, do some blue-skying and come up with a list longer than this. Sure, some of the ideas will be thrown out as worthless. Also sure, some useful criterion measures will emerge.

Let us now turn to the third area that I mentioned in the beginning that is one of the responsibilities of the evaluator. After you have selected the appropriate information tools you must collect and analyze the data. Now it is pretty obvious at this point that I have very little time left to go into a detailed discussion about collection and analysis of data. This is unfortunate in a way because this area is probably as important as or more important than some of the other aspects. However, I would like to limit my remarks to two or theee comments about this area today and let it go at that.

First, the question, when do we collect the data? Keep in mind that the purpose of the evaluation is to aid the decision maker. The information will do the decision maker absolutely no good if it comes in today and the decision was yesterday. This of course means that the data must be collected well in advance of the decision maker's deadline. Hopefully the evaluator and the decision maker will be in constant contact so that the evaluator will know in time when this data must be collected.

Now, what about analyzing the data when he gets it? I'm treating this aspect of the evaluation process very lightly for several reasons. One of course is obvious -- we can't do everything in one brief period. But another reason is, I think sometimes we get so involved with concern about statistical hypothesis tests, analysis of variance, multiple regression analysis, and a few of the other fancier statistical techniques, that we forget to look long and lovingly at the data and try to get it in some kind of a broad perspective. Too often have I seen in comparing an experimental group and a control group the words "no significant difference" with no statement or apparent concern as to the Type Two error involved in the significance test, and where the procedure description indicates it could well be enormous. Too often have I seen lists of items comparing two groups,

with statements "significant at the 5% level" indicated as having some meaning, with no apparent concern that, even if the data were collected and analyzed randomly, one item in twenty would show a significant difference. Too often have I seen a small correlation coefficient indicated as significantly different from zero, yet no mention that it explains practically none of the variance.

Sometimes I think we get so intrigued with the tool that we overlook the mission that the tool is designed to accomplish. Now this may be strange for me to be saying this, as I teach statistics. Yet I'm convinced that in most of the decisions involved in the evaluation process you don't need fancy statistical techniques and when you do need them you can find some expertise on your campus to call upon for assistance. I think that sophisticated statistical techniques are tools to be used in questionable decisions that are close. Quite often they are not close at all -- you can tell broadly from the data what the decision ought to be.

Now that I have probably left the impression that I think statistical techniques are useless, let me hasten to correct that impression. For the decisions that are close, there is no substitute for the correct use of an appropriate statistical tool. We of course cannot discuss detailed statistical methodology today, but let me just recommend to you what I have found to be a useful guide for selecting an appropriate tool I am referring to a table devised by James Beaird of the Oregon Teaching Research Division and available in their CORD Research Training Manual. Beaird requires you to answer four questions, the answers to which, using his guide, lead you to an appropriate statistical tool. The four questions you must answer are:

(1) what is your question --
    do you want to describe, compare, or relate?
(2) how many samples (or variables) do you have?
(3) are your samples independent?
(4) what is the level of your data -- nominal, ordinal or
    better than ordinal?

For further information on this device I refer you to the CORD Manual.

And now just a brief word on the last area of the four that I mentioned that were important in evaluation, that of reporting findings to the decision maker. I am sure that you are well aware that decisions are going to be made whether or not information is presented to the decision maker. It therefore behooves the evaluator to get the information to the decision maker in time for him to make his decision, in a form that he can read and understand, and in such a way that the information presented to him will make sense to him. Careful attention had better be paid to this if we expect him to implement the recommendations suggested by the evaluator. It's very important, it seems to me, that the decision maker be able to scan the evaluator's report in a hurry and pick out the salient points, or chances are he's not going to bother with it. As some one has put it, if one has to search for a needle in a haystack it isn't likely he'll be able to make a stitch in time.

By way of summary of what we have been saying, let me list for you eight steps in the evaluation process as seen by Professor Metfessel of U.S.C.:

(1) Involve the total school community: lay, professional, student
(2) Develop cohesive framework of broad goals and specific objectives
(3) Translate the specific objectives into planned courses of action
(4) Select and/or construct instruments for furnishing measures allowing inferences about program effectiveness
(5) Periodically administer the instrument
(6) Analyze the data collected, using appropriate tools
(7) Interpret the data according to judgmental standards or values
(8) Make recommendations, provide feedback to all involved

Hopefully, the program will be adapted or modified at this point and the cycle of the evaluation process starts once again.

I have been talking today as if all of you out there are evaluators. I suspect that in reality most of you are program directors or involved in

programs in some staff position and that few of you are involved wholeheartedly and completely in the evaluation process. If this is true then I think I have said the right things today, because my objective is to try to get those of you who are involved in the program to be concerned about its evaluation. This concept of evaluation that I've tried to put forth today (and I think it's the current one that most people are accepting) is not one of somebody coming in from the outside looking over your shoulder threatening you, but an evaluation in its true concept should be, as we've described it today, a process whereby information is collected to make better decisions. I hope that you who are involved with planning these programs (and who are not already doing so) will consider allocating someone on your staff and part of your budget to evaluation. Relatively, it's quite inexpensive, and the rewards will be multifold. I hope that you will not fall into the trap of equating program existence with program effectiveness but that you will consider making evaluation part of your program, and assign an evaluator (or maybe a team of evaluators) right from the beginning to initiate the process that we have tried to describe today.

Thank you for the opportunity to participate in your conference.

SOME USEFUL REFERENCES

ALKIN, M. C. (Director)
"Evaluation Comment," periodic publication of the Center for the Study of Evaluation, University of California, Los Angeles

CRAWFORD, J., et al.
CORD (Consortium on Research and Development) National Research Training Manual, 2nd Edition
Teaching Research, Monmouth, Oregon, 1969

HAWKRIDGE, D., et al.
Preparing Evaluation Reports: A Guide for Authors
American Institutes for Research, 135 N. Bellefield Ave., Pittsburgh, Pa., Monograph #6, October, 1970

MAGER, Robert F.
Preparing Instructional Objectives
Fearon Publishers, Palo Alto, Calif., 1962

METFESSEL, N. S. and Michael, W. B.
"Paradigm showing eight major phases of process of evaluating school programs."
U.S.C., mimeographed, 1967

PAULSON, C. F., et al.
A Strategy of Evaluation Design
Teaching Research, Monmouth, Oregon, 1970

PAYNE, S. L.
The Art of Asking Questions
Princeton University Press, 1951

PORTER, Bette
Affective Measures
Teaching Research, Monmouth, Oregon, 1969
(mimeographed)